**REMARKS**

In paragraph 2 of the Action, claims 5 to 14 were allowed. However, claims 1 to 4 were rejected as stated below. Therefore, the applicants have filed request for continued examination. In the amendments, claims 5 to 14 have not been amended. Therefore, claims 5 to 14 are still in condition of allowance.

In paragraph 7 of the Action, claims 1 and 2 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art, in view of *Otsuka* (US Patent No. 6,546,367).

In paragraph 8 of the Action, claims 3 and 4 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art, in view of *Otsuka*, further in view of *Vermeulen et al.* (US Patent No. 6,810,379).

The Applicants respectfully traverse the rejections and request reconsideration. In view of the rejections cited in paragraphs 7 and 8, claims 1 and 3 have been amended to clarify the features of the invention and add a new limitation. With the amendments, claims 1 to 4 are not unpatentable over applicant's admitted prior art, in view of the cited references, for the reasons explained below.

As recited in claim 1, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string;

a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition by referring to the voice segment dictionary.

Further, the method comprises the steps of providing the prosody generation module with a phoneme duration determination unit that includes both a duration rule table containing empirically found phoneme durations and a duration prediction table containing phoneme durations predicted by statistical analysis; designating an utterance speed; selecting one of the duration rule table and the duration prediction table according to the utterance speed; and determining a phoneme duration by using, when the utterance speed exceeds a threshold contained in the duration rule table, said duration rule table and, when said utterance speed does not exceeded the threshold, said duration prediction table.

In particular, the method includes the steps of designating the utterance speed and selecting one of the duration rule table and the duration prediction table according to the utterance speed. Accordingly, it is possible to designate various utterance speeds according to nature of speech, and to determine the phoneme duration based on whether the utterance speed exceeds the threshold.

*Otsuka* discloses a speech synthesizing method and apparatus as well as a storage medium for setting a phoneme duration for a phoneme string to achieve a specified speech-production time and provide a natural phoneme duration regardless of a length of speech production time. In *Otsuka*, Fig. 2 shows a block diagram of a flow structure of the speech synthesizing apparatus. In Fig. 2, a phoneme duration setting unit 5 sets a phoneme duration in accordance with control data, representing speech production speed stored in a control data storage unit 2. According to *Otsuka*,

using the phoneme duration value, the phoneme duration is determined according to the equation (3a). When the obtained phoneme duration is smaller than a threshold value, the phoneme duration is determined according to the equation (3b), in which the phoneme duration is equal to the threshold value, so that reproduced speech becomes natural (col. 3, line 16 to col. 4, line 60).

In particular, in *Otsuka*, the phoneme duration di for each phoneme αi of the phoneme string is determined such that the phoneme string constructed by phonemes αi ($1 \leq i \leq N$) in the phoneme duration setting section is phonated within the speech production time T, determined based on the control data representing speech production speed stored in the control data storage unit 2 (col. 3, line 63 to col. 4, line 2).

In *Otsuka*, FIG. 5 is a flowchart showing the process of determining a phoneme duration according to the first embodiment, which shows the detailed process of steps S5 and S6 in FIG. 3. In step S107, the phoneme duration di for the phoneme αi is determined so as to coincide with the speech production time T of the expiratory paragraph, based on the phoneme duration initial value for all the phonemes in the expiratory paragraph obtained in the previous process and the standard deviation of the phoneme αi (i.e., determined according to the equation (3a)). If the phoneme duration di obtained in step S107 is smaller than a threshold value θαi set for the phoneme αi, the threshold value θαi is set to di (steps S108 and S109) (col. 6, lines 1-10).

In the invention recited in claim 1, the method includes the steps of designating the utterance speed and selecting one of the duration rule table and the duration prediction table according to

the utterance speed. Accordingly, it is possible to designate various utterance speeds according to nature of speech, and to determine the phoneme duration based on whether the utterance speed exceeds the threshold.

In Otsuka, the speech production time T is determined based on the control data representing speech production speed stored in the control data storage unit 2. There is no disclosure or suggestion regarding the steps of designating the utterance speed and selecting one of the duration rule table and the duration prediction table according to the utterance speed.

Therefore, *Otsuka* does not disclose nor suggest the features of the invention recited in claim 1. Further, even though *Otsuka* is combined with Applicant's admitted prior art, the invention recited in claim 1 is not obvious.

As recited in claim 3, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition while referring to the voice segment dictionary.

Further, the method comprises the steps of providing the prosody generation module with a pitch contour determination unit that has both a rule table empirically obtained and a prediction table predicted by statistical analysis; designating an utterance speed; selecting one of the rule table and the prediction table

according to the utterance speed; and determining a pitch contour by using accent and phrase components contained in, when the utterance speed exceeds a threshold contained in the rule table, the rule table and, when the utterance speed does not exceeded the threshold, the prediction table.

In particular, the method includes the steps of designating the utterance speed and selecting one of the rule table and the prediction table according to the utterance speed. Accordingly, it is possible to designate various utterance speeds according to nature of speech, and to determine the pitch contour based on whether the utterance speed exceeds the threshold.

As explained above, in Otsuka, there is no disclosure or suggestion regarding the steps of designating the utterance speed and selecting one of the rule table and the prediction table according to the utterance speed.

Therefore, *Otsuka* does not disclose nor suggest the features of the invention recited in claim 3. Further, even though *Otsuka* is combined with Applicant's admitted prior art, the invention recited in claim 3 is not obvious.

*Vermeulen et al.* has disclosed a client/server architecture for text-to-speech synthesis. In Fig. 1 in *Vermeulen et al.*, a text-to-speech system 10 is provided with a prosody generation unit 16. The prosody generation unit 16 produces timing and pitch information for speech synthesis. According to *Vermeulen et al.*, the pitch is determined from a rule set or statistical model (col. 2, line 1 to line 21).

In the invention, the pitch contour determination unit determines a pitch contour by determining both accent and phrase

components with the rule table when a user-designated utterance speed exceeds a threshold contained in the rule table, and with the prediction table when the utterance speed does not exceed the threshold.

In *Vermeulen et al.*, it is simply stated that the pitch is determined from a rule set or statistical model. There is no disclosure or suggestion regarding the method of setting the pitch contour base on the threshold as claimed in the invention.

Therefore, *Vermeulen et al.* do not disclose or suggest the features of the invention recited in claim 3. Even though *Otsuka* and *Vermeulen et al.* are combined with Applicant's admitted prior art, the invention recited in claim 3 is not obvious.

As recited in claim 15, a method of the invention controls high-speed reading in a text-to-speech conversion system. The method comprises: inputting a text into the text-to-speech conversion system; generating a phoneme and prosody character string of the text with a text analysis module; creating a duration rule table containing a first phoneme duration obtained empirically; creating a duration prediction table containing a second phoneme duration obtained through statistical analysis; designating an utterance speed; comparing the utterance speed with a threshold value; selecting one of the duration rule table and the duration prediction table according to the utterance speed; determining a third phoneme duration with a phoneme duration determination unit according to the one of the duration rule table and the duration prediction table; generating a synthesis parameter of at least a voice segment, the third phoneme duration, and a fundamental frequency of the phoneme and prosody character string with a prosody generation module; and generating a synthetic waveform through waveform superimposition with a speech generation

module according to the synthesis parameter and a voice segment dictionary containing a voice segment as a basic source of voice.

In particular, the method comprises the steps of designating an utterance speed; comparing the utterance speed with a threshold value; selecting one of the duration rule table and the duration prediction table according to the utterance speed; and determining a third phoneme duration with a phoneme duration determination unit according to the one of the duration rule table and the duration prediction table.

As explained above, in *Otsuka*, there is no disclosure or suggestion regarding the steps of designating the utterance speed and selecting one of the rule table and the prediction table according to the utterance speed. Therefore, *Otsuka* does not disclose nor suggest the features of the invention recited in claim 15.

As recited in claim 18, a method of the invention controls high-speed reading in a text-to-speech conversion system. The method comprises: inputting a text into the text-to-speech conversion system; generating a phoneme and prosody character string of the text with a text analysis module; creating a duration rule table containing a first phoneme duration obtained empirically; creating a duration prediction table containing a second phoneme duration obtained through statistical analysis; designating an utterance speed; comparing the utterance speed with a threshold value; selecting one of the duration rule table and the duration prediction table according to the utterance speed; determining a third phoneme duration with a phoneme duration determination unit according to the one of the duration rule table and the duration prediction table; generating a synthesis parameter

of at least a voice segment, the third phoneme duration, and a fundamental frequency of the phoneme and prosody character string with a prosody generation module; and generating a synthetic waveform through waveform superimposition with a speech generation module according to the synthesis parameter and a voice segment dictionary containing a voice segment as a basic source of voice.

In particular, the method comprises the steps of designating an utterance speed; comparing the utterance speed with a threshold value; and selecting one of the duration rule table and the duration prediction table according to the utterance speed.

As explained above, in *Otsuka*, there is no disclosure or suggestion regarding the steps of designating the utterance speed, and selecting one of the rule table and the prediction table according to the utterance speed. Therefore, *Otsuka* does not disclose nor suggest the features of the invention recited in claim 18.

As explained above, the cited references do not disclose or suggest all of the features of the invention recited in claims 1, 3, 15 and 18. Further, even though the cited references are combined with Applicant's admitted prior art, the invention is not obvious. Therefore, the invention is not patentable over the applicant's admitted prior art in view of the cited references.

Reconsideration and allowance are earnestly solicited.

One-month extension of time is requested. The credit card payment form in the amount of $1,320 (RCE filing fee $790, one-month extension fee $130, and fee for two additional independent claims $400) has been attached herewith.

Respectfully submitted,

Kazunao Kubotera
Reg. No. 51,194

TAKEUCHI & KUBOTERA, LLP
200 Daingerfield Rd.
Suite 202
Alexandria, VA 22314
Tel. (703) 684-9777
Fax. (703) 684-1390